

Deep Learning for Sequences – Assignment 2 – Part 3

Daniel Juravski 206082323
Eyal Orbach 015369317

Our best parameters for both POS and NER:

Embedding vector size = 50

Ngram = 5

Hidden layer size = 50

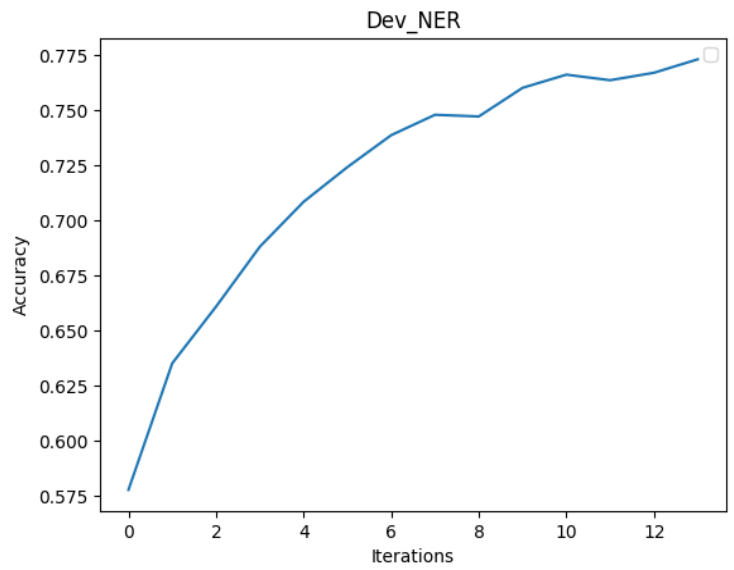
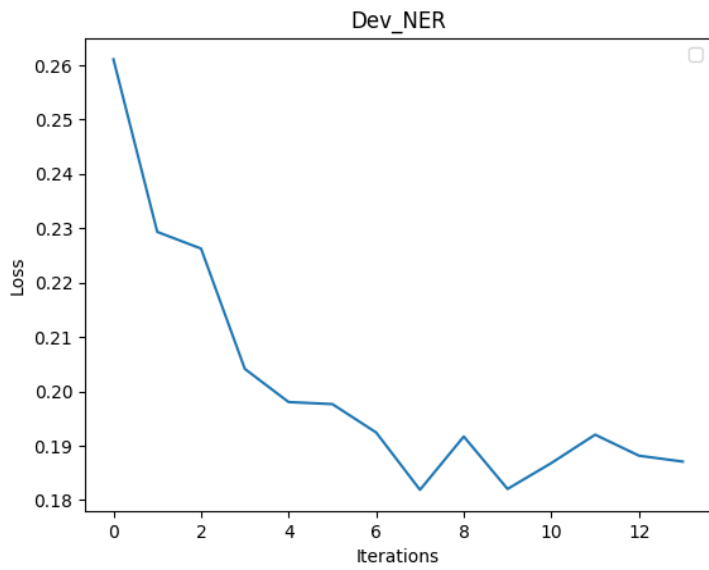
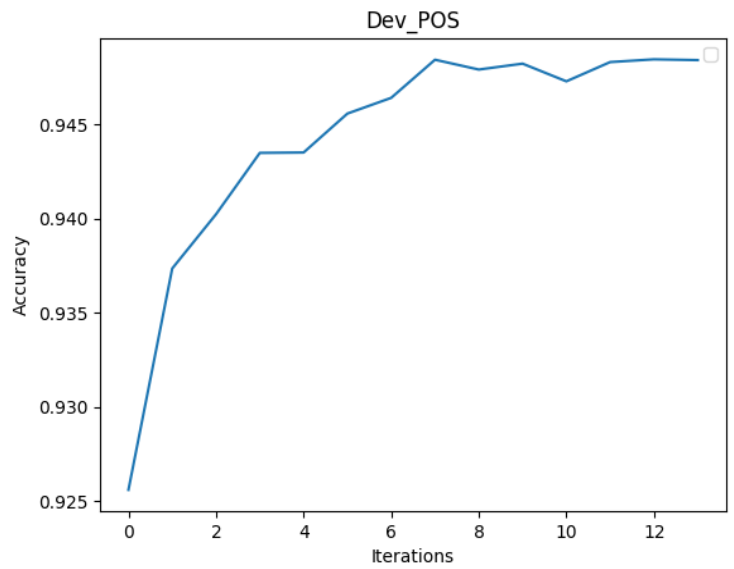
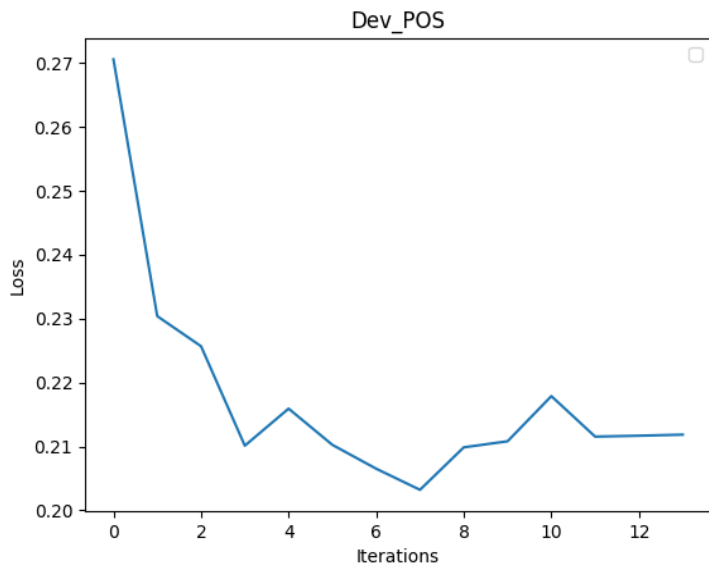
Optimizer = “Adam”

Learning Rate = 0.002

Epochs = 7

Considerations:

- There were words that appear in the training file but not the embeddings file, we treated them in a similar way like in part 1, by adding four different vectors to unknown words and mapping to the correct unknown vector by the word's features (the same feature as at part 1).
- The embedding vocabulary represented all with lower-case. So appropriately we searched for the lowercased representation of the word, if it was not found then we mapped it to the correct unknown word vector.
- After several run-ups, we saw that the pre-trained vectors doesn't improve our accuracy by a great difference, but the model is convergent faster for a good result.



Notes:

1. Our model has learned in 7 epochs. We wanted to see the dev evaluation loss and accuracy some more times so we evaluate it twice in an epoch. Once in the middle of the epoch, and once at the end of the epoch.
2. The NER accuracy graph, calculated with the metric of ignoring the 'O' tag.